Halo in customer satisfaction measures

The role of purpose of rating, number of attributes and customer involvement

Jochen Wirtz

NUS Business School, National University of Singapore, Singapore

Abstract *Firms usually measure customer satisfaction on an attribute-by-attribute basis in order to identify and improve potential weaknesses, and to fortify their strengths in service delivery. However, research has shown that halo can threaten the interpretability of such data. Also, halo is particularly acute in satisfaction measurement of services with a high degree of ambiguous and credence attributes. This paper examines three halo-reducing methods developed in psychology and organizational behavior in the context of customer satisfaction. The perceived purpose of evaluation (evaluative vs developmental) and the number of attributes measured (few vs many) were examined in an experimental design, and the level of product involvement (low vs high) was examined using a quasi-experimental design. The data showed reduced halo when the respondents were presented with a developmental rather than evaluative purpose, when more rather than fewer attributes were measured, and when subjects were highly involved with the service.*

Introduction

Customer satisfaction is increasingly recognized as a main ingredient for success in the market place (Weiser, 1995). A growing body of literature shows that customer satisfaction leads to repeat purchase, loyalty, positive word-of-mouth and increased long term profitability (e.g. Heskett *et al.*, 1994). Hence, many firms are tracking their customers' satisfaction, guided by the dictum "what is not measured is not managed" (Srinivasan, 1996; Wirtz and Tomlin, 2000). Many do so, on an attribute-by-attribute level, in order to identify and improve potential weaknesses and to fortify their strengths.

However, recent studies demonstrated that halo can undermine the interpretability of attribute-specific satisfaction data, obscure the identification of the strengths and weaknesses, and make attribute-specific comparisons across competing brands and products unreliable (Wirtz and Bateson, 1995; Wirtz, 2000). For instance, misinterpretation of halo-contaminated data can cause firms to single out the wrong attributes as weaknesses and make sub-optimal investments into improving them, or to overlook weak attributes that may be exploited by competition. This topic is of particular importance for services, as they often have attributes that are difficult to evaluate. Credence

attributes refer to attributes which consumers may not be able to evaluate reliably even after consumption (Mattila and Wirtz, 2002). For example, it is virtually impossible for a consumer to assess the quality of attributes relevant for a root canal procedure (Alford and Sherrell, 1996) or in many professional services (Patterson *et al.*, 1997). Furthermore, ambiguous attributes may be evaluated by the consumer, but can be interpreted in different ways. For example, a messy desk in a travel agency may be a sign of poor service quality, or a signal of a busy and successfully managed enterprise. Because credence and ambiguous attributes are difficult to evaluate, the perception of both types of attributes may be influenced by the performance of the more easily assessable search and experience attributes, resulting in halo effects (Wirtz, 2000). In other words, satisfaction measurement in a service context may be more susceptible to halo than measurement in a goods context. It therefore seems to be important to understand how to control and reduce halo in service satisfaction measurement. The realization of the importance of halo in general, and the importance of halo in a services context in particular, led to the advance of a number of propositions on how halo can be reduced in attribute-specific satisfaction measures (Wirtz, 1996). Some of these propositions were subsequently empirically tested (Wirtz, 2001). The objectives of this study are to further extend this line of research on halo reduction in satisfaction data.

## Definition and causes of halo
Definition of halo
Halo has been studied particularly in the fields of psychology (e.g. Hauenstein and Alexander, 1991), social psychology (e.g. Nisbett and Wilson, 1977) and organizational behavior (e.g. Fox *et al.*, 1994) ever since it was discovered by Wells (1907) and christened by Thorndike (1920). Despite the halo literature being confounded by a number of conceptual definitions (Balzer and Sulsky, 1992), an agreement on the technical nature of halo has been established (Murphy *et al.*, 1993). Halo is defined as the excess correlation over and above the true correlation between attributes (Murphy and Jako, 1989). Halo assimilates the evaluation of different attributes, flattens the overall profile of evaluations, and compresses the differences among attribute evaluations (Murphy *et al.*, 1993).

Causes of halo
Fisicaro and Lance (1990) identified three conceptually distinct causes of halo, which they presented in the form of causal models. These models are:

(1) the general impression model;

(2) the salient dimension model; and

(3) the inadequate discrimination model.

A recently proposed associonist model has been incorporated into Fiscario and Lance's framework by Wirtz (1996) as shown in Figure 1.

G

R1   R2   R1 ———▶ R2   R1   R2

↑   ↑   ↑   ↑   ↑   ↑

T1   T2   T1   T2   T1   T2

**A.   General
        Impression
        Model**

**B1.   Salient
        Dimension
        Model**

**C.   Inadequate
        Discrimination
        Model**

**B2.   Associonist
        Effects Model**
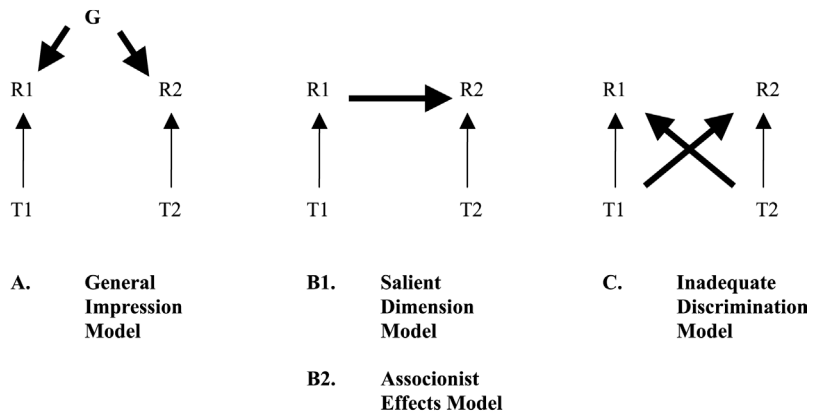
Key:

G	:	a rater's general impression
T1 and T2	:	a rater's true attribute satisfaction level
R1 and R2	:	a rater's reported attribute satisfaction level

Figure 1.
Causal models of
halo effects

Note: Disturbance terms are omitted for parsimony

Source: Adapted from Fisicaro and Lance (1990) and Wirtz (1996)

General impression model. This model suggests that raters have a tendency to think of something in general as being good or inferior, and thereby coloring their judgment of its separate dimensions with this feeling or evaluation (Thorndike, 1920). Nisbett and Wilson (1977) also suggested that a global evaluation may impact the evaluation of individual attributes. For instance, a customer who likes a particular brand may tend to view all the product's attributes as excellent, even though some may in fact only show average performance. Referring to Figure 1 (A), a rater's general impression G, in this case of the brand, has a causal effect on the dimensional evaluations R1 and R2 (the various attributes). The final result is an inflated correlation between the attributes R1 and R2.

Salient dimension model. This model posits that one or more salient dimensions can influence the evaluation of performance on less salient dimensions (Kozwolski *et al.*, 1986). Robbins' (1989, p. 444) definition of the halo error concurred with this model:

> The tendency for an evaluator to let the assessment of a product in one attribute influence his evaluation of that product on other attributes.

For instance, a customer may allow his evaluation of a waiter's friendliness (a potentially salient attribute) influence his evaluation of the waiter's uniform (a potentially less important attribute). As shown in Figure 1 (B1), the more salient attribute R1 (in this case, the assessment of the waiter's friendliness) directly impacts the less salient attribute R2 (the assessment of the waiter's uniform).

Associonist model. This model proposes that making a judgment about one attribute can prime or activate similarly toned information, which subsequently affects the evaluation of other attributes (Judd *et al.*, 1991; Tourangeau and Rasinski, 1988). Hence, the sequence in which attributes are being evaluated can cause halo (Sinclair, 1988). This points to the significance of contextual and order effects in the study of halo. For instance, assume a customer had a bad check-in experience. If that customer then evaluated the check-in experience before other elements of the airline's service such as inflight meals, inflight entertainment and cabin crew service, his cognitive processing would have been primed to the negative aspects of the experience, causing him to be negatively biased (cf. Feldman, 1981). In a services context, Ruyter *et al.* (1997a) showed that the sequence of encounters has an impact on overall satisfaction evaluations, indicating carry over effects akin to halo.

Inadequate discrimination model. In this model, halo is caused by a rater's failure to distinguish between different dimensions of performance, which are distinct and potentially independent (Saal *et al.*, 1980, p. 415). Balzer and Sulsky (1992) also referred to this bias as the dimensional similarity halo, while Ghiselli and Brown (1948) called it a logical error. Raters tend to use inferred relations among performance dimensions as the basis of a set of evaluations, rather than to depend on actual observation of the individual attribute performances (Balzer and Sulsky, 1992). This may be due to the rater's inability (Cooper, 1981a, b) or unwillingness (Banks and Murphy, 1985) to discriminate among attributes, or it may reflect a rater's subconscious attempts to maintain cognitive consistency (Holbrook, 1983). To illustrate, a salesperson's professional appearance may be perceived by his clients as an indication of his product knowledge. However, the attributes "professional appearance" and "product knowledge" are independent of one another. In this case, the customer's rating on the salesperson's "professional appearance" may influence his rating on "product knowledge". This can be seen in Figure 1 (C), where ratings on T1 and T2 affect evaluations R2 and R1, respectively.

Hypotheses development
After having explored various causes of halo, this section advances hypotheses of three methods that are proposed to reduce halo in satisfaction data. These three methods were selected as they seem to be effective in disciplines other than consumer behavior, and they could potentially be used in applied satisfaction research. The three methods are:

(1) the perceived purpose of evaluation;

(2) the number of attributes measured; and

(3) the level of product involvement of the respondent.

Purpose of evaluation
Banks and Murphy (1985) suggested that raters may be unwilling to discriminate among different attributes as a result of interpersonal and

affective processes. In the context of customer satisfaction, a customer may be less willing to discriminate among attributes when the evaluation is perceived as being part of a performance appraisal. For instance, feedback to a branch or hospital department may be perceived as evaluative, where feedback (especially negative feedback) may have implications for the staff that had served the customer. Customers may not want to feel responsible for an organization's internal evaluation of those who served them. On the other hand, customers may be more willing to evaluate, if the exercise is perceived as developmental in nature (e.g. for the development of better staff training programs). They may also be more willing to exert more cognitive effort, if they believe their feedback will lead to better products in the future. It is therefore proposed that a satisfaction study with a perceived developmental nature will result in lower halo than one that is perceived as evaluative in nature. Also, Wirtz (1996) suggested in his conceptual paper that emphasizing a developmental purpose reduces halo caused by general impression and inadequate discrimination.

H1. Halo is lower when the perceived purpose of a study is developmental rather than evaluative in nature.

Number of attributes
Murphy *et al.* (1993, p. 222) proposed that:

> Halo errors seem most likely when there are only a few dimensions, each of which is highly relevant to one's overall evaluation, and less likely when there are many dimensions, several of which are apparently unrelated to overall performance.

By increasing the number of attributes to be rated, the rater is forced to expend more cognitive effort than when there are only a few attributes. As a result, the rater is able to better distinguish between attribute performances, and thereby reduces halo caused by inadequate discrimination during evaluation. Wirtz (2001) provided indicative support for this proposition in a field setting demonstrating a reduction in inter-item correlation. However, this reduction in observed correlation could not be unambiguously attributed to a reduction in halo (i.e. rival hypotheses could not be ruled out due to a survey method being used in his study). Here, it is aimed to retest the field findings using a true experimental design.

H2. Evaluation of a large number of attributes results in lower halo than evaluation of a small number of attributes.

Level of involvement
In psychology, involvement has been suggested to determine a respondent's willingness to discriminate among dimensions (Banks and Murphy, 1985). The literature suggests that the more a person is involved with a product, the more likely he/she is to expend more cognitive effort in its purchase, consumption and evaluation processes (Hague and Flick, 1989). It seems reasonable to assume that this argument can be extended to attribute satisfaction measures.

For example, Curren and Harich (1994) showed that when an evaluation is seen as relatively unimportant (i.e. consumers have low involvement with the product or the rating), people will simply transfer their own affective feelings to a product's attributes, which then results in halo. Similarly, Mattila (1998) has shown that customers with low involvement tend to follow heuristics in satisfaction evaluations and that their evaluations are more easily influenced by moods than those of more involved respondents. Increasing involvement levels appears to address all causes of halo, because of the rater's high personal relevance, which results in the rater's commitment to higher cognitive effort.

   H3. The ratings of respondents with high product involvement show lower halo than those of respondents with low involvement.

Method
Experimental design
Wirtz (2001) used a field setting in a service context for his study on halo reduction methods. The study compared observed inter-item correlation coefficients between experimental conditions and drew the conclusions based on the observed differences. However, the true correlations between the various attributes used in his study were not known. Hence, it was impossible to determine how much halo was reduced, or whether true correlations other than halo had been removed as well. To circumvent this shortcoming, the present study differs in two important ways. First, halo was induced directly into the data set. Using an experimental design, it was possible to manipulate the performance of one attribute at two levels, while holding all other attributes constant. As a result, there should be zero correlation between the satisfaction measures of the manipulated and the non-manipulated attributes. Any observed correlation between them would be caused by halo. As such, the amount of halo reduced via the proposed methods could be determined.

   Second, Wirtz (2001) had included attribute-specific perceived performance and disconfirmation-of-expectations measures in addition to attribute-specific satisfaction scales in his study. In the present study, perceived performance and disconfirmation-of-expectations measures were excluded for the following three reasons. First, by responding to performance and disconfirmation scales, demand effects might be created. By repeatedly going through a series of scale items, subjects may become more aware of certain attributes. It may also heighten their desire for consistent ratings across the attribute-specific measures of performance, disconfirmation-of-expectations, and satisfaction. This effect can be avoided by excluding the former two measures from the questionnaire altogether. Second, a questionnaire that requires attribute-by-attribute evaluations of performance, disconfirmation-of-expectations and satisfaction can easily become tedious and repetitious for the subjects. This may dampen respondents' willingness to evaluate their consumption experience, and/or introduce fatigue into the evaluative process (Laurent, 1972), leading to higher halo. Finally, applied satisfaction research typically does not measure expectations, performance and disconfirmation to reduce costs, to

allow for other managerially more important issues to be included in the survey, and/or to keep the length of the survey manageable. Thus, in the light of these three considerations, only attribute-specific satisfaction measures were included in the present study.

Research setting and stimuli design
A videotape was used in this study. Bateson and Hui (1992) demonstrated the high ecological validity of videotapes, defined as "the applicability of the results of a laboratory analogues to non-laboratory, real life settings" (McKechnie, 1977). A video presentation standardizes the service experience for all subjects, and discourages subjects from bringing their personal expectations into the experiment. Both reduce random variation in the data and increase the power of hypothesis testing. Furthermore, the video allows time to be compressed and a service environment to be simulated. The video used in this study was filmed at a fast food restaurant, which depicted a customer's service encounter from the time he entered, to the time he received his food and found a seat.

Manipulation of treatment variables
A $2 \times 2 \times 2$ factorial design was employed. The performance manipulation of the attribute "cleanliness of restroom" (excellent and neutral performance) was used to induce halo into the data set. This attribute was chosen based on in-depth interviews with fast food restaurant managers and patrons. Furthermore, this attribute seemed suitable for halo induction, as it is unrelated to many other dimensions of a fast food service experience.

Two other manipulations were used to test the hypotheses advanced in this paper, namely, the purpose of evaluation (evaluative and developmental), and the number of attributes to be evaluated (few and many). These independent variables were operationalized through the research design by using different versions of the questionnaire. The level of involvement was examined using a quasi-experimental design in which involvement was measured and not directly manipulated.

Induction of halo. The performance of "cleanliness of restroom" was manipulated using written scenarios describing the condition of the restroom at the restaurant. Two scenarios (excellent and neutral performance) were used to induce halo in the experiment (Figure 2).

Manipulation of purpose of evaluation. The purpose of evaluation was operationalized at two levels: evaluative and developmental. A short paragraph in the introduction letter explained the purpose of the evaluation (Figure 3).

Manipulation of few/many attribute measurement manipulations. To determine the most relevant attributes to measure, an exploratory study involving interviews with fast food restaurant managers and customers was conducted. An initial list of 15 attributes was obtained and then reduced to the ten most important attributes. The manipulation of number of attributes was operationalized at two levels: few (five attributes) and many (ten attributes) (Table I).

Excellent Performance Condition:

When Mr. X visited the washroom before leaving the restaurant, he noted that the floor was surprisingly clean and dry. The toilet paper was also well stocked, and because the flush was functioning effectively, there wasn't the unpleasant odor that he had half expected to come with the typical public washroom. All in all, Mr. X thought it was a well-maintained restroom.

Average Performance Condition:

When Mr. X visited the washroom before leaving the restaurant, he was neither impressed nor distressed at the restroom's cleanliness, as it was about average.

Figure 2.
Manipulation of
attribute for halo
induction

Developmental Condition:

Your response will be used to help fast food restaurants to improve their service quality. It will also be used to develop better training programmes for employees in fast food restaurants.

Evaluative Condition:

Your response will be used to help fast food restaurant managers monitor and evaluate their branch performance. It will also facilitate their task in identifying employees who have performed below par.

Figure 3.
Manipulation of purpose
of evaluation

Attributes	Symbols	Attributes included in the many/few attribute conditions	
		Many (ten attributes)	Few (five) attributes
Cleanliness of restroom (manipulated attribute)	Restrm	✓	✓
Waiting time	Time	✓	✓
Helpfulness of waiter	Help	✓	✓
Availability of seats	Seats	✓	✓
Range of selection of menu	Menu	✓	✓
Courtesy of waiter	Court	✓	
Cleanliness of restaurant	Clean	✓	
Design of restaurant	Design	✓	
Price level	Price	✓	
Tidiness of waiter	Tidy	✓	

Table I.
Manipulation of
"many" and "few"
attribute conditions

Measures

A commonly used one-item seven-point semantic differential scale was used to measure the satisfaction level with each of the attributes (Westbrook and Oliver, 1981), anchored in 1 = "extremely satisfied" and 7 = "extremely dissatisfied".

To measure involvement, the revised product involvement inventory (RPII) was used (McQuarrie and Munson, 1992). RPII is an adaptation of Zaichkowsky's (1985) product involvement inventory (PII). In comparison to the PII, the RPII is shorter (ten instead of 20 items), but still shows high criterion validity. The anchors of the seven-point semantic differential RPII scale used in this study were:

(1) "fast food is important" and "is unimportant";

(2) "is not relevant to me" and "is relevant to me";

(3) "means a lot to me" and "means nothing to me";

(4) "is exciting" and "is unexciting";

(5) "is uncool" and "is cool";

(6) "matters to me" and "does not matter to me";

(7) "is boring" and "is interesting";

(8) "is fun" and "is not fun";

(9) "is appealing" and "is unappealing"; and

(10) "is of no concern to me" and "is of concern to me".

The questionnaire was pretested, whereby the pretest subjects were asked to "think aloud" while completing the questionnaire during the pretests. Any item that was found to be unclear or perceived differently than intended was noted. Two items were found to be ambiguous in this research context. The "relevant-irrelevant" item was not understood, and the "dull-neat" item was thought to mean "untidy-tidy". In a subsequent revision, these items were changed to "relevant to me – not relevant to me" and "uncool-cool" as shown in the list of final items used in the previous paragraph. In a second pre-test, the revised scale had a high reliability with a Cronbach alpha of 0.90.

Experimental procedure

Experimental sessions were conducted with a total of 316 subjects, consisting of 99 undergraduate and 217 junior college students. In each session, an oral introduction was given to familiarize the subjects with the procedures. Following that, the video was screened after which the subjects were asked to complete the questionnaire. The time needed for each session was less than 20 minutes.

Data analysis
Manipulation check
A manipulation check showed that the performance manipulation of the attribute "cleanliness of restroom" was successful. A one-way ANOVA demonstrated a main effect of the manipulation on Restrm ($F(1;314) = 1,921$; $p < 0.001$). The mean scores for Restrm show that the manipulation was in the intended direction, with ratings of 1.48 and 3.78 for the excellent and neutral conditions, respectively. Involvement was measured rather than manipulated. It was the only multi-item scale in this study and displayed good reliability ($\alpha = 0.90$).

Next, two sets of analyses were performed. First, it was necessary to establish that halo had indeed been successfully induced into the dataset. Second, the three hypotheses were tested.

Testing for presence of halo
In this study, the attribute "cleanliness of restroom" was manipulated and all other attributes were held constant via the use of a video presentation. As a result, the true correlation between Restrm with manipulated performance levels and the satisfaction scores of the non-manipulated attributes were designed to be zero. Using these experimental procedures, a significant correlation between the manipulated attribute and other attributes would indicate the existence of halo.

Table II shows the correlations between all attribute-specific satisfaction measures. It can be seen that Restrm was significantly correlated to seven of the nine non-manipulated attributes:

(1) courtesy of waiter (Court);

(2) cleanliness of restaurant (Clean);

(3) design of restaurant (Design);

(4) helpfulness of waiter (Help);

(5) range of menu (Menu);

(6) availability of seats (Seats); and

(7) waiting time (Time).

These correlation coefficients ranged from $r = 0.11$ to 0.42, $p < 0.05$, suggesting that halo was successfully induced into the data.

Furthermore, the mean satisfaction scores for all attributes were compared between the excellent and average cleanliness of restroom manipulations. All satisfaction mean scores of the non-manipulated attributes were higher in the excellent than in the neutral Restrm condition, and seven of the nine mean pairs were significantly different at $p < 0.05$, again suggesting that halo was successfully induced into the dataset (Table III).

Attributes	Mean	STD	1	2	3	4	5	6	7	8	9	10
1 Restrm	2.75	1.29	*1.00*									
2 Time	2.26	1.07	*0.30**	1.00								
3 Help	2.13	0.97	*0.26**	0.51	1.00							
4 Seats	1.92	0.98	*0.17**	0.48	0.47	1.00						
5 Menu	2.84	1.14	*0.15**	0.30	0.40	0.29	1.00					
6 Court	1.99	0.82	*0.19**	0.29	0.75	0.42	0.30	1.00				
7 Clean	2.55	1.00	*0.42**	0.35	0.47	0.24	0.07	0.41	1.00			
8 Design	3.13	0.89	*0.21**	0.29	0.34	0.20	0.28	0.36	0.43	1.00		
9 Price	4.35	1.35	*0.11*	0.22	0.15	0.11	0.40	0.10	0.11	0.30	1.00	
10 Tidy	2.34	0.94	*0.11*	0.22	0.51	0.30	0.27	0.65	0.40	0.29	0.20	1.00

Notes: *Significant level at $p < 0.05$; the italic numbers are the correlation coefficients between the manipulated Restrm attribute and all other not manipulated attributes

	Restrm	Time	Help	Seats	Menu	Court	Clean	Design	Price	Tidy
Excellent condition	1.48	1.99	1.88	1.74	2.70	1.81	2.18	3.00	4.24	2.23
Neutral condition	3.78	2.52	2.35	2.08	2.96	2.16	2.88	3.25	4.46	2.44
Delta means	2.30	0.53	0.47	0.34	0.26	0.35	0.70	0.25	0.22	0.27
p-value	<0.001	<0.001	<0.001	<0.001	0.019	0.003	<0.001	0.043	0.158	0.077

Notes: Restrm was the manipulated attribute, which was used to induce halo into the other attribute ratings; 1 = "extremely satisfied"; 7 = "extremely dissatisfied"

Table III.
Mean scores of average
and excellent
performance conditions

Developmental versus evaluative purpose (H1)
In line with Murphy and Balzer's (1989) suggestion, the hypotheses were tested by comparing the correlation coefficients of Restrm and the satisfaction measures of the non-manipulated attributes across the experimental conditions. To test for significance between coefficient pairs, the individual inter-item correlation coefficients were transformed to z-scores using Fisher's r-to-z transformation. It converts individual inter-item correlation coefficients into normalized distributions.

H1 advanced that halo is lower when respondents perceived the satisfaction study to be of a developmental rather than evaluative nature. Eight of the nine correlation pairs moved in the predicted direction, five of which were significant or marginally significant at the 0.05 or 0.10 levels, respectively (Table IV). The remaining correlation coefficient (Clean) remained the same across the two experimental conditions. Furthermore, the overall average inter-item correlation was significantly lower for the developmental condition ($r = 0.16$) than for the evaluative condition ($r = 0.25$; $p < 0.05$), supporting *H1*.

Next, we explored whether the difference in mean attribute ratings between the two restroom conditions would be less in the developmental than in the evaluative condition (Table V). We found that five of the nine attribute pairs were shifted significantly less ($p < 0.05$) or marginally significantly less ($p < 0.10$) in the developmental condition than in the evaluative condition. Also, the overall mean shift was significantly less in the developmental condition (delta mean = 0.31) than in the evaluative condition (delta mean = 0.45; $p < 0.012$). In summary, the correlation and shift in mean analyses supported *H1*.

	Correlation coefficients between not manipulated attributes and the manipulated restroom attribute for two experimental conditions	
Not manipulated attributes	Evaluative condition ($n = 158$)	Developmental condition ($n = 158$)
Time	0.34	0.25*
Help	0.27	0.24
Seats	0.24	0.08*
Menu	0.19	0.10*
Court	0.23	0.15
Clean	0.42	0.42
Design	0.28	0.12*
Price	0.13	0.09
Tidy	0.19	0.03**
Average inter-item correlation	0.25	0.16**

Table IV.
Comparison of inter-item correlations (*H1*)

Notes: The items Court, Clean, Design, Price, and Tidy were available for only half the sample, as they were only included in the few attribute condition; * significant difference in hypothesized direction at $p < 0.10$; ** significant difference in hypothesized direction at $p < 0.05$

Manipulation	Restrm performance manipulation	Mean satisfaction ratings of not manipulated attributes									
		Time	Help	Seats	Menu	Court	Clean	Design	Price	Tidy	Total
Evaluative condition											
1 Excellent	1.42	1.86	1.79	1.65	2.63	1.63	2.17	2.86	4.17	2.03	2.20
2 Neutral	3.90	2.40	2.23	2.12	2.99	2.00	2.86	3.26	4.55	2.39	2.65
3 (2-1) Delta means	2.48	0.54	0.44	0.47	0.36	0.37	0.69	0.40	0.38	0.36	0.45
4 *p*-value	<0.001	<0.001	0.001	0.002	0.029	0.019	0.001	0.031	0.124	0.055	<0.001
Developmental condition											
5 Excellent	1.55	2.12	1.97	1.82	2.77	1.97	2.19	3.14	4.30	2.41	2.42
6 Neutral	3.85	2.63	2.46	2.05	2.94	2.30	2.88	3.23	4.37	2.49	2.73
7 (6-5) Delta means	2.30	0.51	0.49	0.23	0.17	0.33	0.69	0.09	0.07	0.08	0.31
8 *p*-value	<0.001	0.002	0.001	0.070	0.162	0.038	0.001	0.305	0.401	0.344	<0.001
9 (3-7) Difference in mean shift between evaluative and developmental conditions		0.03	−0.05	0.24	0.19	0.04	0.00	0.31	0.31	0.28	0.14
10 *p*-value	0.401	0.401	0.677	0.016	0.070	0.379	1.00	0.016	0.080	0.034	0.012

Notes: The means in the total column exclude the manipulated restroom attribute; 1 = "extremely satisfied"; 7 = "extremely dissatisfied"

Table V.
Comparison of means
across experimental
conditions (*H1*)

Number of attributes (H2)
H2 advanced that the greater the number of attributes to be evaluated, the lower is the level of halo in attribute-specific satisfaction measures. Table VI shows the correlation coefficient pairs of Restrm and four non-manipulated attributes, which were used in both experimental conditions (Time, Help, Seats, and Menu). Three out of the four coefficient pairs moved in the predicted direction, and the average inter-item correlation was significantly lower in the many ($r = 0.18$) than in the few condition ($r = 0.25$, $p < 0.05$), supporting *H2*.

The differences in mean attribute ratings were in the expected direction for all attributes, i.e. they were less in the many condition than in the few condition (Table VII). Also, two of the four attribute pairs were shifted significantly less ($p < 0.05$) in the many than in the few condition. Finally, the overall mean shift was significantly less in the many (delta mean = 0.32) than in the few condition (delta mean = 0.49; $p = 0.004$). In summary, the correlation and shift in mean analyses supported *H2*.

Involvement (H3)
H3 proposed that the higher involvement a person has with the product, the lower is the halo effect experienced. Involvement was not manipulated but tested in a quasi-experimental design. The median score of involvement was used to split the data into a low and a high involvement group. The median score was 4.10.

As shown in Table VIII, all correlation pairs moved in the predicted direction, and seven of the nine correlations were significant at $p < 0.05$ or marginally significant at $p < 0.10$. In addition, the average inter-correlation was significantly lower for the high involvement group ($r = 0.13$) than for the low involvement group (0.29; $p < 0.05$).

The differences in mean attribute ratings were in the expected direction for all attributes but one (Tidy, whose delta means were the same in both conditions; Table IX). Six of the nine-delta means were significantly different at

	Correlation coefficients between not manipulated attributes and the manipulated restroom attribute for two experimental conditions	
Not manipulated attributes	Few attribute condition ($n = 163$)	Many attribute condition ($n = 153$)
Time	0.29	0.31
Help	0.30	0.21*
Seats	0.19	0.14
Menu	0.23	0.05**
Average inter-item correlation	0.25	0.18**

Notes: Correlations could only be tested for those attributes that were included in both the many and few attribute conditions; *significant difference in hypothesized direction at $p < 0.10$; **significant difference in hypothesized direction at $p < 0.05$

Table VI.
Comparison of inter-item correlation (*H2*)

| Manipulation | Restrm attribute manipulations | Means in satisfaction ratings of not manipulated attributes | | | | |
		Time	Help	Seats	Menu	Total
Few attribute condition						
1	Excellent	2.03	1.87	1.74	2.44	2.02
2	Neutral	2.59	2.42	2.12	2.89	2.51
3 (2-1)	Delta means	0.56	0.55	0.38	0.45	0.49
4	*p*-value	0.001	<0.001	0.007	0.007	<0.001
Many attribute condition						
5	Excellent	1.94	1.89	1.73	2.97	2.13
6	Neutral	2.44	2.27	2.05	3.04	2.45
7 (6-5)	Delta means	0.50	0.38	0.32	0.07	0.32
8	*p*-value	0.001	0.004	0.024	0.354	<0.001
9 (3-7)	Difference in mean shift between the few and many attribute conditions	0.06	0.17	0.06	0.38	0.17
10	*p*-value	0.310	0.061	0.294	0.002	0.004

column 2 (Restrm attribute manipulations) values:

Restrm attribute manipulations
1.47
3.85
2.38
<0.001
1.50
3.90
2.40
<0.001

Notes: Only attributes that were included in both the many and few attribute conditions are included in this table. The means in the total column exclude the manipulated restroom attribute; 1 = "extremely satisfied"; 7 = "extremely dissatisfied"

Table VII.
Comparison of means across experimental conditions (*H2*)

Not manipulated attributes	Correlation coefficients between not manipulated attributes and the manipulated restroom attribute for two experimental conditions	
	Low involvement ($n = 154$)	High involvement ($n = 162$)
Time	0.41	0.15**
Help	0.30	0.19**
Seats	0.23	0.05**
Menu	0.18	0.11*
Court	0.20	0.18
Clean	0.55	0.27**
Design	0.28	0.14*
Price	0.27	0.01*
Tidy	0.20	0.05
Average inter-item correlation	0.29	0.13**

Table VIII.
Comparison of inter-item correlation (*H3*)

Notes: The items Court, Clean, Design, Price, and Tidy were available for only half the sample, as they were only included in the few attribute condition; * significant difference in hypothesized direction at $p < 0.10$; ** significant difference in hypothesized direction at $p < 0.05$

the 0.05 level, and one was marginally significant at 0.10. Furthermore, the overall mean shift was significantly less in the high involvement group (delta mean = 0.16) than in the low involvement group (delta mean = 0.56; $p < 0.001$). In summary, the correlation and shift in mean analyses supported *H3*.

Discussion and directions for further research
Summary of findings
First, the study showed that a perceived developmental purpose led to lower halo than an evaluative purpose. It appeared that when the subjects believed that their input could help the service firm to improve and deliver better service in the future, they were more willing to participate in the evaluation, discriminate among the attributes, and rely less on general impressions.

Second, earlier survey findings suggest that increasing the number of attributes reduces halo (Wirtz, 2001). These findings were replicated here in an experimental setting. Subjects seemed to expend more cognitive effort in order to discriminate between the attributes when there were more rather than less attributes, and thereby decreased halo caused by inadequate discrimination (cf. Murphy *et al.*, 1993). The research setting of this study differs from the study conducted by Wirtz (2001) in many ways. For instance, the employee profiling service used in Wirtz's (2001) study was likely to be a one-off experience for most users, while the fast food context of this study is a commonplace service. Most users of the employee profiling service were armed with no prior knowledge of the consumption process, while the script for an encounter at the fast food restaurant is very familiar to most customers. Despite these methodological and contextual differences, using a larger number of attributes

Manipulation	Restrm performance manipulation	Time	Help	Mean satisfaction ratings of not manipulated attributes							
				Seats	Menu	Court	Clean	Design	Price	Tidy	Total
Low involvement condition											
1 Excellent	1.56	1.97	1.88	1.80	2.65	1.77	2.10	2.84	3.87	2.10	2.24
2 Neutral	3.92	2.77	2.53	2.29	2.99	2.14	3.12	3.33	4.47	2.24	2.80
3 (2-1) Delta means	2.36	0.80	0.65	0.51	0.34	0.37	1.02	0.49	0.60	0.14	0.56
4 p-value	<0.001	<0.001	<0.001	0.001	0.028	0.029	<0.001	0.010	0.030	0.072	<0.001
High involvement condition											
5 Excellent	1.42	2.01	1.88	1.68	2.75	1.83	2.24	3.12	4.51	2.32	2.37
6 Neutral	3.82	2.22	2.13	1.83	2.92	2.16	2.59	3.14	4.46	2.46	2.53
7 (6-5) Delta means	2.40	0.21	0.25	0.15	0.17	0.33	0.35	0.02	0.05	0.14	0.16
8 p-value	<0.001	0.101	0.040	0.163	0.185	0.035	0.059	0.474	0.433	0.248	0.018
9 (3-7) Difference in mean shift between the low and high involvement conditions		0.59	0.40	0.36	0.17	0.04	0.67	0.47	0.55	0.00	0.40
10 p-value	<0.001	<0.001	<0.001	0.001	0.094	0.382	<0.001	0.001	0.007	1.00	<0.001

Notes: The means in total column exclude the manipulated restroom attribute; 1 = "extremely satisfied"; 7 = "extremely dissatisfied"

Table IX.
Comparison of means
across experimental
conditions

had effectively reduced halo in both studies, indicating generalization beyond the immediate research contexts.

Third, the study showed that high product involvement led to lower halo. With higher personal relevance, subjects appeared motivated to engage in more cognitive effort in the evaluation processes (e.g. Hague and Flick, 1989). Looking at the findings (i.e. magnitude in reduction of correlations and delta mean shifts), involvement perhaps showed the most powerful reduction in halo of the three methods tested.

Managerial implications

The results of this study offer a few important managerial implications. First, the results showed that by stressing the developmental purpose of satisfaction studies, subjects seem more willing to discriminate among the attributes. Managers can use statements such as "we want to serve you better", or "help us improve" as used by Ikea on its customer feedback forms to communicate a developmental purpose.

Second, managers should consider using more attributes in their satisfaction studies. Not only do they get more accurate data by reducing halo, they also attain more detailed information on the various aspects of their products. For instance, in this study in the many attribute condition (i.e. ten attributes) the data were considerably "cleaner", and a more comprehensive picture of the restaurant's service could be obtained than in the few attribute condition (five attributes).

Finally, there is evidence that high involvement reduces halo. Perceived involvement or importance of the product to respondents can be included in applied satisfaction surveys. By controlling for involvement and by examining the attribute ratings of highly and lowly involved customers separately, managers can get a better picture of their strengths and weaknesses. Our findings suggest that the ratings of the highly involved customers are more accurate and show lower halo than ratings provided by less involved customers.

It is easy to incorporate these methods into most research designs. Hence, it is advised that managers adopt these measures where suitable. This way, satisfaction measurement can be improved by reducing potential halo errors.

Directions for future research

This study may be replicated and extended in a number of ways. First, the halo induced in this study was lower than desired. The manipulation of the attribute "cleanliness of restroom" may not have induced strong halo, perhaps because the manipulation was not sufficiently strong. It was manipulated at two levels: "excellent" and "neutral performance". A "poor" performance condition might have triggered much stronger halo. Also, more than one attribute could be manipulated, as it has been demonstrated in an earlier study that the manipulation of two attributes in the same direction causes significantly more halo than the manipulation of only one attribute (Wirtz, 2000).

The use of "many" attributes as a halo reduction method was shown to be effective in Wirtz (2001) as well as the present study. Both used ten and five attributes to operationalize "many" and "few" attributes, respectively. However, neither study quantified the limits of "many" and "few". In any evaluation, exhaustion and fatigue set in beyond a certain number of attributes (Laurent, 1972; Helgeson and Ursic, 1994). Subsequently, biases such as central tendency, consistency, leniency, and halo can interact to inflate the observed correlation. Therefore, an inverse U-shaped relationship between number of attributes and level of halo seems likely. It is thus useful to learn more about the range of optimal number of attributes and the boundary conditions that may influence this range.

This study continues a stream of research examining halo reduction methods borrowed from the fields of psychology, social psychology and organizational behavior, and applied them in a customer satisfaction context. The advance in halo research achieved in this present study represents only a small step towards a better understanding of halo. There are other interesting phenomena expounded in the psychology and organizational behavior literatures that can provide potentially useful and interesting insights when examined from a marketing perspective. For example, some respondents may be more prone to halo than others. This assumption underlies much of the research on the cognitive dynamics of halo (e.g. Cooper, 1981b; Feldman, 1981; Nathan and Lord, 1983), and the search for individual and/or situational differences that might explain halo (Murphy and Jako, 1989). In psychology, halo effects have been shown to be smaller when respondents are familiar rather than unfamiliar with what is being evaluated. Respondents with little familiarity may have relatively undifferentiated impressions, whereas experienced respondents may develop more fine-grained evaluations (Murphy et al., 1993). The same logic seems to apply in the context of marketing. For example, only discerning consumers can evaluate the various attributes of a fine cognac, a sports car or a symphony concert. Here, inexperienced consumers may base their attribute-specific evaluations on general impressions or on the few attributes they are familiar with. Therefore, one may propose that halo errors are lower when respondents are experienced with the consumption of the product/service to be evaluated rather then when they are not. Future research can put this hypothesis to test.

One cause of general impressions halo may be affective overtones driving the assessment of individual dimensions (Wirtz and Bateson, 1995). Affect is increasingly seen as an important determinant of satisfaction (e.g. Mano and Oliver, 1993; Oliver, 1993; Wirtz et al., 2000; Yu and Dean, 2001), and Holbrook (1983) developed a potentially interesting method for controlling halo caused by affective overtones. He proposed a structural modeling approach using feedback loops from overall affect back to the individual attribute evaluations and thereby identifying the portion of attribute specific evaluation that is caused by affective overtones. Ruyter et al. (1997b) showed that there are carry-over effects in satisfaction ratings between different steps in a service process,

and again, this type of halo could potentially be modeled using structural equation modelling.

Finally, improved modelling and using better measures may reduce halo (cf. Wirtz and Mattila, 2001). Wirtz and Lee (2003) showed that the quality of measures improves with the number of items, whereby a six-item 7-point semantic differential satisfaction scale (e.g. Oliver and Swan 1989) performed best in their study, followed by a four-item 7-point semantic differential scale (e.g. Eroglu and Machleit 1990), and then followed by a single-item 11-point percentage scale (e.g. Westbrook 1980). Also, the 11-point single-item scale performed better than all other single-item 7-point scales. These findings seem to suggest that multi-item scales achieve more finely grained measurement by tapping into satisfaction from different angles than single-item scales. Furthermore, it seems that a more finely grained rating on a 11-point scale with extreme scale anchors (not at all satisfied – completely satisfied) is able to capture satisfaction better than 7-point single-item satisfaction scales. Wirtz and Lee's (2003) study examined the quality of measures in the context of overall satisfaction measurement, and it would be interesting to explore the applicability of these findings in the context of halo in attribute-specific measures. For example, does a 11-point scale show less halo than a 5-point scale, or does a three-item scale per attribute show less halo than the single-item measures tested in this present study? These are interesting avenues for future work.

Summary

The results from this study show that it is possible to use halo-reduction methods developed in the fields of psychology, social psychology and organizational behavior in a customer satisfaction context. In particular, the study supported the use of three halo reduction methods:

(1) positioning the purpose of the study as developmental in nature rather than evaluative;

(2) asking respondents to evaluate more rather than fewer attributes; and

(3) controlling for involvement in attribute-specific satisfaction measures, whereby more involved respondents showed less halo that less involved ones.

References

Alford, B.L. and Sherrell, D.L. (1996), "The role of affect in consumer satisfaction judgments of credence-based services", *Journal of Business Research*, Vol. 37, pp. 71-84.

Andrews, F.M. and Withey, S.B. (1976), *Social Indicators of Well-Being*, Plenum Press, New York, NY.

Balzer, W. and Sulsky, L.M. (1992), "Halo and performance appraisal research: a critical examination", *Journal of Applied Psychology*, Vol. 77 No. 6, pp. 975-85.

Banks, C.G. and Murphy, K.R. (1985), "Toward narrowing the research-practice gap in performance appraisal", *Personnel Psychology*, Vol. 38, pp. 335-45.

Bateson, J.E.G. and Hui, M.K. (1992), "The ecological validity of photographic slides and videotapes in simulating the service setting", *Journal of Consumer Research*, Vol. 19, pp. 271-81.

Cooper, W.H. (1981a), "Conceptual similarity as a source of illusory halo in job performance ratings", *Journal of Applied Psychology*, Vol. 66, pp. 302-7.

Cooper, W.H. (1981b), "Ubiquitous halo", *Psychological Bulletin*, Vol. 90, pp. 218-44.

Curren, M.T. and Harich K.R. (1994), "Consumers' mood states: the migrating influence of personal relevance on product evaluations", *Psychology and Marketing*, Vol. 11 No. 2, pp. 91-107.

Eroglu, S.A. and Machleit, K.A. (1990), "An empirical study of retail crowding: antecedents and consequences", *Journal of Retailing*, Vol. 66, Summer, pp. 201-21.

Feldman, J.M. (1981), "Beyond attribution theory: cognitive processes in performance appraisal", *Journal of Applied Psychology*, Vol. 66, pp. 127-48.

Fisicaro, S.A. and Lance, C.E. (1990), "Implications of three causal models for the measurement of halo error", *Applied Psychological Measurement*, Vol. 14, pp. 419-29.

Fox, S., Caspy, T. and Reisler, A. (1994), "Variables affecting leniency, halo and validity of self-appraisal", *Journal of Occupational and Organizational Psychology*, Vol. 67, pp. 45-56.

Ghiselli, E.E. and Brown, C.W. (1948), *Personnel and Industrial Psychology*, 1st ed., McGraw-Hill, New York, NY.

Hague, R.A. and Flick, L.F. (1989), "Enduring involvement: conceptual and measurement issues", *Advances in Consumer Research*, Vol. 16, pp. 690-6, Association for Consumer Research, Provo, UT.

Hauenstein, N.M.A. and Alexander, R.A. (1991), "Rating ability in performance judgments: the joint influence of implicit theories and intelligence", *Organizational Behavior and Human Decision Processes*, Vol. 50, pp. 300-23.

Helgeson, J.G. and Ursic, M.L. (1994), "The role of affective and cognitive decision-making processes during questionnaire completion", *Psychology and Marketing*, Vol. 11 No. 5, pp. 493-510.

Heskett, J.L., James, T.O., Loveman, G.W., Sasser, W.E. Jr and Schlesinger, L.A. (1994), "Putting the service-profit chain to work", *Harvard Business Review*, Vol. 72 No. 2, pp. 164-74.

Holbrook, M.B. (1983), "Using a structural model of halo effect to assess perceptual distortion due to affective overtones", *Journal of Consumer Research*, Vol. 10, pp. 247-52.

Judd, C.M., Drake, R.A., Dowing, J.W. and Krosnick, J.A. (1991), "Some dynamic properties of attitude structures: context-induced response facilitation and polarization", *Journal of Personality and Social Psychology*, Vol. 60, pp. 193-202.

Kozlowski, S.W.J., Kirsch, M.P. and Chao, G.T. (1986), "Job knowledge, ratee familiarity, conceptual similarity and halo error: an exploration", *Journal of Applied Psychology*, Vol. 71, pp. 45-9.

Laurent, A. (1972), "Effects of length on reporting behavior in survey interviews", *Journal of the American Statistical Association*, pp. 278-305.

McKechnie, G.E. (1977), "Simulation techniques in environment psychology", in Stokols, D. (Ed.), *Perspectives on Environment and Behavior: Theory, Research and Applications*, Plenum, New York, NY, pp. 169-89.

McQuarrie, E.F. and Munson, J.M. (1992), "A revised product involvement inventory: improved usability and validity", *Advances in Consumer Research*, Vol. 19, pp. 108-15.

Mano, H. and Oliver, R.L. (1993). "Assessing the dimensionality and structure of the consumption experience: evaluation, feeling and satisfaction", *Journal of Consumer Research*, Vol. 20, pp. 451-66.

Mattila, A. (1998), "An examination of consumer heuristic cues in making satisfaction judgments", *Psychology & Marketing*, Vol. 15 No. 5, pp. 477-501.

Mattila, A. and Wirtz, J. (2002), "The impact of knowledge types on the consumer search process – an investigation in the context of credence services" *International Journal of Service Industry Management*, Vol. 13 No. 3, pp. 214-30.

Murphy, K.R. and Balzer, W.K. (1989), "Rater errors and rating accuracy," *Journal of Applied Psychology*, Vol. 74 No. 4, pp. 619-24.

Murphy, K.R. and Jako, R. (1989), "Under what circumstances are observed intercorrelations greater or smaller than true intercorrelations?", *Journal of Applied Psychology*, Vol. 74, pp. 827-30.

Murphy, K.R., Jako, R. and Anhalt, R.L. (1993), "Nature and consequences of halo error: a critical analysis", *Journal of Applied Psychology*, Vol. 78, pp. 218-25.

Nisbett, R.E. and Wilson, T.D. (1977), "The halo effect: evidence for unconscious alteration of judgements", *Journal of Personality and Social Psychology*, Vol. 35 No. 4, pp. 250-6.

Nathan, B.R. and Lord, R.G. (1983), "Cognitive categorization and dimensional schemata: a process approach to the study of halo in performance ratings", *Journal of Applied Psychology*, Vol. 71, pp. 29-32.

Oliver, R.L. (1993), "Cognitive, affective and attribute bases of the satisfaction response", *Journal of Consumer Research*, Vol. 20, pp. 418-30.

Oliver, R.L. and Bearden, W.O. (1983), "The role of involvement in satisfaction processes", *Advances in Consumer Research*, Vol. 11, pp. 250-5.

Oliver, R.L. and Swan J.E. (1989), "Consumer perceptions of interpersonal equity and satisfaction in transactions: a field survey approach", *Journal of Marketing*, Vol. 53, April, pp. 21-35.

Patterson, P.G., Johnson, L.W. and Spreng, R.A. (1997), "Modeling the determinants of customer satisfaction for business-to-business professional services", *Journal of the Academy of Marketing Science*, Vol. 25 No. 1, pp. 4-17.

Robbins, S.P. (1989), *Organisational Behaviour*, 4th ed., Prentice Hall, Englewood Cliffs, NJ.

Ruyter, K., Wetzels, M., Lemmink, J. and Mattson, J. (1997a), "The dynamics of the service delivery process: a value-based approach", *International Journal of Research in Marketing*, Vol. 14, pp. 231-43.

Ruyter, K., Lemmink, J., Wetzels, M. and Mattson, J. (1997b), "Carry-over effects in the formation of satisfaction: the role of value in a hotel service delivery process", *Advances in Services Marketing and Management*, Vol. 6, JAI Press Inc., Greenwich, CT, pp. 61-77.

Saal, F.E., Downey, R.G. and Lahey, M.A. (1980), "Rating the ratings: assessing the psychometric quality of rating data", *Psychological Bulletin*, Vol. 88, pp. 413-28.

Sinclair, R.C. (1988), "Mood, categorization, breadth, and performance appraisal: the effects of order of information acquisition and affective state on halo, accuracy, information retrieval and evaluations", *Organizational Behavior and Human Decision Processes*, Vol. 42, pp. 22-46.

Srinivasan, M. (1996), "Marketing impact rests on resource allocation", *Marketing News*, Vol. 30 No. 6, p. 4.

Thorndike, E.L. (1920), "A constant error in psychological ratings", *Journal of Applied Psychology*, Vol. 4, pp. 25-29.

Tourangeau, R. and Rasinski, K.A. (1988), "Cognitive processed underlying context effects in attitude measurement", *Psychological Bulletin*, Vol. 103, pp. 299-314.

Westbrook, R.A. (1980), "A rating scale for measuring product/service satisfaction", *Journal of Marketing,* Vol. 44, Fall, pp. 68-72.

Westbrook, R.A. and Oliver, R.L. (1981), "Developing better measures of consumer satisfaction: some preliminary results", *Advances in Consumer Research*, Vol. 8, pp. 94-9.

Weiser, C.R. (1995), "Championing the customer", *Harvard Business Review*, Vol. 73 No. 6, pp. 113-16.

Wells, F. (1907), "A statistical study of literary merit", *Archives of Psychology*, Vol. 1 No. 7.

Wirtz, J. (1996), "Controlling halo in attribute-specific customer satisfaction measures – towards a conceptual framework", *Asian Journal of Marketing*, Vol. 5 No. 1, pp. 41-58.

Wirtz, J. (2000), "An examination of the presence, magnitude and impact of halo on customer satisfaction measures", *Journal of Retailing and Consumer Services*, Vol. 7, pp. 89-99.

Wirtz, J. (2001), "Improving the measurement of customer satisfaction: a test of three methods to reduce halo", *Managing Service Quality*, Vol. 11 No. 2, pp. 99-111.

Wirtz, J. and Bateson, J.E.G. (1995), "An experimental investigation of halo effects in satisfaction measures of service attributes", *International Journal of Service Industry Management*, Vol. 6 No. 3, pp. 84-102.

Wirtz, J. and Lee, M.C. (2003), "An examination of the quality and context-specific applicability of commonly used customer satisfaction measures", *Journal of Services Research*, Vol. 5 No. 4.

Wirtz, J. and Mattila, A.S. (2001), "Exploring the role of alternative perceived performance measures and needs-congruency in the consumer satisfaction process", *Journal of Consumer Psychology*, Vol. 11 No. 3, pp. 181-92.

Wirtz, J. and Tomlin, M. (2000), "Institutionalizing customer-driven learning through fully integrated customer feedback systems", *Managing Service Quality*, Vol. 10 No. 4, pp. 205-15.

Wirtz, J., Mattila, A.S. and Tan, R.L.P. (2000), "The moderating role of target-arousal state on the impact of affect on satisfaction – an examination in the context of service experiences", *Journal of Retailing*, Vol. 76 No. 3, pp. 347-65.

Yu, Y.T. and Dean, A. (2001), "The contribution of emotional satisfaction to consumer loyalty", *International Journal of Service Industry Management*, Vol. 12 No. 3, pp. 234-50.

Zaichkowsky, J.L. (1985), "Measuring the involvement construct", *Journal of Consumer Research*, Vol. 12, pp. 341-51.